On the Characterization of Q-Superlinear Convergence
of Quasi-Newton Interior-Point Methods
for Nonlinear Programming

H.J. Martinez

Z. Parada

R.A. Tapia

Report Documentation Page		Form Approved OMB No. 0704-0188

Public reporting burden for the collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden, to Washington Headquarters Services, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington VA 22202-4302. Respondents should be aware that notwithstanding any other provision of law, no person shall be subject to a penalty for failing to comply with a collection of information if it does not display a currently valid OMB control number.

1. REPORT DATE **APR 1995**	2. REPORT TYPE	3. DATES COVERED **00-00-1995 to 00-00-1995**
4. TITLE AND SUBTITLE **On the Characterization of Q-Superlinear Convergence of Quasi-Newton Interior-Point Methods for Nonlinear Programming**		5a. CONTRACT NUMBER
		5b. GRANT NUMBER
		5c. PROGRAM ELEMENT NUMBER
6. AUTHOR(S)		5d. PROJECT NUMBER
		5e. TASK NUMBER
		5f. WORK UNIT NUMBER
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) **Computational and Applied Mathematics Department ,Rice University,6100 Main Street MS 134,Houston,TX,77005-1892**		8. PERFORMING ORGANIZATION REPORT NUMBER
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)		10. SPONSOR/MONITOR'S ACRONYM(S)
		11. SPONSOR/MONITOR'S REPORT NUMBER(S)
12. DISTRIBUTION/AVAILABILITY STATEMENT **Approved for public release; distribution unlimited**		
13. SUPPLEMENTARY NOTES		
14. ABSTRACT		
15. SUBJECT TERMS		

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES **16**	19a. NAME OF RESPONSIBLE PERSON
a. REPORT **unclassified**	b. ABSTRACT **unclassified**	c. THIS PAGE **unclassified**			

Standard Form 298 (Rev. 8-98)
Prescribed by ANSI Std Z39-18

On the Characterization of Q-Superlinear Convergence of Quasi-Newton Interior-Point Methods for Nonlinear Programming.*

H. J. Martinez[†]   Z. Parada[‡]   and R. A. Tapia[§]

February, 1994

Abstract

In this paper we extend the well-known Boggs-Tolle-Wang characterization of Q-superlinear convergence for quasi-Newton methods for equality constrained optimization to quasi-Newton interior-point methods for nonlinear programming. Critical issues in this extension include, the choice of the centering parameter, the choice of the steplength parameter, and the determination of the primary variables.

Keywords: Interior-point methods, primal-dual variables, nonlinear programming, superlinear convergence, Boggs-Tolle-Wang characterization, Dennis-Moré characterization.

Abbreviated Title: Characterization of superlinear convergence.

AMS(MOS) subject classifications: 65K, 49M, 90C.

1

1 Introduction.

In 1974 Dennis and Moré [3] gave a characterization of those quasi-Newton methods for the nonlinear equation problem which produce iterates which are Q-superlinearly convergent. This characterization immediately carries over to unconstrained optimization by working with the nonlinear equation (gradient equal to zero) that results from the first-order necessary conditions. Similarly the Dennis-Moré characterization can be carried over to equality constrained optimization by working with the nonlinear system corresponding to the first-order necessary conditions. This nonlinear system, involves the two groups of variables (x, y). Here x is the vector of primal variables, and y is the vector of dual variables corresponding to the equality constraints. Hence the approach characterizes Q-superlinear convergence in terms of the variable-pair (x, y). Indeed, the first authors to establish Q-superlinear convergence for various secant methods for equality constrained optimization, Han [8] in 1976, Tapia [12] in 1977, and Glad [7] in 1979, did so using this approach and established Q-superlinear convergence in the pair (x, y). Not long after, in 1982, Boggs, Tolle, and Wang [1] observed that under certain assumptions, various quasi-Newton secant methods for equality constrained optimization actually give Q-superlinear convergence in the primal variable x alone. They then proceeded to establish a characterization of those quasi-Newton methods that produced iterates which are Q-superlinearly convergent in the primal variable x alone. Nocedal and Overton [9] in 1983, and Fontecilla, Steihaug, and Tapia [6] in 1987 derived the Bogss-Tolle-Wang characterization under less restrictive assumptions than those used by Boggs, Tolle, and Wang. Finally, in 1987 Stoer and Tapia [11] gave a very short and self-contained derivation of the Boggs-Tolle-Wang characterization.

Recently, there has been activity in extending the successful primal-dual Newton interior-point method from linear programming to general nonlinear programming. In linear programming, the primal-dual Newton method, although not initially presented in this manner, is now recognized as a damped and perturbed Newton method applied to the Karush-Kuhn-Tucker (KKT) necessary conditions. This interpretation serves as the vehicle for their extension to nonlinear programming. In 1992, El-Bakry, Tapia, Tsuchiya, and Zhang [5] established the local convergence properties of the Newton interior-point method for NLP . These convergence results are in line with those of the standard Newton's method. In 1993 Yamashita and Yabe [14] considered quasi-Newton interior-point formulations and used the Dennis-Moré theory to derive a characterization of those methods which gave Q-superlinear convergence in all of the variables. The KKT conditions involve a vector of primal variables x, a vector of dual variables y corresponding to equality constraints, and a vector of dual variables z corresponding to the nonnegative constraints on the primal variables x. Hence the variables consist of the triple (x, y, z), and x and z are required to be nonnegative.

We see from the Boggs-Tolle-Wang theory that while the variables involved in quasi-Newton methods for equality constrained optimization are the primal variables x and the

2

dual variables y, it is possible to obtain a characterization result in terms of the primal variables alone. Hence, in some sense the primal variables are also the primary variables. This understanding led us initially to try to obtain a characterization in terms of the primal variables x also for quasi-Newton interior-point methods. However, we could not do so without including some undesirable assumption on the interaction between the primal variable x and the dual variables z. This, in turn, led us to search for a characterization in terms of the variables (x, z). Our search was successful and is the subject of the current research. It is interesting then, that in the sense alluded to above, the primary variables for quasi-Newton interior-point methods are the nonnegative variables (x, z).

This paper is organized as follows. In Section 2, with an eye towards our main characterization result, we study the characterization of Q-superlinear convergence for a damped and perturbed quasi-Newton method for the nonlinear equation problem. Our intention is not to give a complete theory for the topic, but to develop the tools needed for our interior-point application. In Section 3, we describe our quasi-Newton interior-point method. In Section 4, we derive an equivalence between our quasi-Newton interior-point method and a damped and perturbed quasi-Newton method for a system of nonlinear equations that involves only the variables (x, z). This equivalence has the flavor of the approach taken by Stoer and Tapia [11] when they derived the Boggs-Tolle-Wang characterization. In Section 5, we apply the theory developed in Section 2 to the equivalent formulation obtained in Section 4 and establish our main characterization results.

2 Characterization for damped and perturbed quasi-Newton methods.

In this section we formulate and study a damped and perturbed quasi-Newton method for the nonlinear equation problem. Our objective is to derive characterization results concerning Q-superlinear convergence that can be used to establish our main characterization theorem for quasi-Newton interior-point methods in Section 5.

Consider the nonlinear equation problem

$$F(x) = 0 \tag{2.1}$$

where $F : \mathbf{R}^n \to \mathbf{R}^n$. Recall that the standard Newton's method theory assumptions for problem (2.1) are

S1. There exists $x^* \in \mathbf{R}^n$ such that $F(x^*) = 0$.

S2. The Jacobian matrix $F'(x^*)$ is nonsingular.

S3. The Jacobian operator F' is Lipschitz continuous at x^* in an open convex neighborhood of x^*, with Lipschitz constant $\gamma > 0$.

As usual the expressions F_k, F_{k+1} and F_* denote the evaluation of the function F at the points x_k, x_{k+1}, and x^* respectively. Similar notation will be used for other quantities.

By a **damped and perturbed quasi-Newton method** for problem (2.1), we mean the construction of the iteration sequence

$$x_{k+1} = x_k - \alpha_k A_k^{-1}[F_k - r_k] \quad , \quad k = 0, 1, 2, \ldots \quad . \tag{2.2}$$

In (2.2), $0 < \alpha_k \leq 1$ is the steplength parameter, $r_k \in \mathbf{R}^n$ is a perturbation vector, and A_k is an approximation to F'_k.

We begin by collecting some known useful facts. Toward this end let $e_k = x_k - x^*$ and $s_k = x_{k+1} - x_k$; assume **S1 - S3**, and that $\{x_k\}$ converges to x^*.
There exists a constant $\rho > 0$ such that for k sufficiently large

$$\frac{1}{\rho}\|e_k\| \leq \|F_k\| \leq \rho\|e_k\| \ . \tag{2.3}$$

A proof of (2.3) can be found, for example, in Dembo, Eisenstat, and Steihaug [2]. It follows that

$$\frac{\|e_{k+1}\|}{\|e_k\|} \to 0 \Rightarrow \frac{\|s_k\|}{\|e_k\|} \to 1 \ , \tag{2.4}$$

and

$$\frac{\|e_{k+1}\|}{\|e_k\|} \to 0 \Leftrightarrow \frac{\|F_{k+1}\|}{\|s_k\|} \to 0 \quad . \tag{2.5}$$

To establish (2.4) we merely need to observe that $e_{k+1} = s_k + e_k$. Moreover, (2.5) follows directly once we write

$$\frac{\|F_{k+1}\|}{\|e_k\|} = \frac{\|F_{k+1}\|}{\|s_k\|}\frac{\|s_k\|}{\|e_k\|} \quad .$$

The next two theorems will motivate choices for the steplength α_k and the perturbation vector r_k.

Theorem 2.1. *Let $\{x_k\}$ be generated by (2.2). Assume that **S1**, **S2**, and **S3** hold and that $x_k \to x^*$. Then any two of the following statements imply the third:*

(i) $x_k \to x^*$ *Q-superlinearly.*

(ii) $\lim_{k\to\infty} \frac{\|\alpha_k r_k + (1-\alpha_k)F_k\|}{\|s_k\|} = 0.$

4

(iii) $\lim_{k\to\infty} \frac{\|(A_k - F'_*)s_k\|}{\|s_k\|} = 0.$

Proof. Adding and substracting the appropriate quantities, we have

$$F_{k+1} = [F_{k+1} - F_k - F'_* s_k] - [A_k - F''_*]s_k + [\alpha_k r_k + (1 - \alpha_k)F_k]. \tag{2.6}$$

From (2.5), (i) is equivalent to

$$\lim_{k\to\infty} \frac{\|F_{k+1}\|}{\|s_k\|} = 0.$$

Using Lemma 4.1.15 in [4] we have

$$\|F_{k+1} - F_k - F'_* s_k\| \le \frac{\gamma \|s_k\|}{2}(\|e_{k+1}\| + \|e_k\|). \tag{2.7}$$

The remainder of the proof is fairly straightforward.

□

Observe that if for all k, $\alpha_k = 1$ and $r_k = 0$, then (2.2) becomes the standard quasi-Newton method; moreover, in this case condition (ii) is trivially satisfied and Theorem 2.1 reduces to the standard Dennis-Moré characterization.

Condition (ii) tells us that essentially for Q-superlinear convergence we must have $\alpha_k \to 1$ and $r_k = o(\|s_k\|)$. We are somewhat concerned with this latter requirement for the following reason. Our expectation is to be able to control the size of the perturbation vector r_k; however, at the begining of the iteration when we must choose r_k, the step s_k is unknown to us. For this reason we look for a similar condition involving $\|F_k\|$, a quantity which is readily available. However, we must add an assumption concerning the rate of convergence of $\{x_k\}$.

Theorem 2.2. *Let $\{x_k\}$ be generated by (2.2). Assume that* **S1, S2,** *and* **S3** *hold and that* $x_k \to x^*.$ *Then any two of the following statements imply the third.*

$(i)'$ $x_k \to x^*$ *Q-superlinearly.*

$(ii)'$ $\lim_{k\to\infty} \frac{\|\alpha_k r_k + (1-\alpha_k)F_k\|}{\|F_k\|} = 0$ *and the convergence of $\{x_k\}$ to x^* is Q-linear.*

$(iii)'$ $\lim_{k\to\infty} \frac{\|(A_k - F'_*)s_k\|}{\|s_k\|} = 0.$

Proof. We must show that any two conditions in Theorem 2.1 are equivalent to the corresponding two conditions in Theorem 2.2. Observe that from (2.3), the fact that $s_k = e_{k+1} - e_k$, and the Q-linear convergence of $\{x_k\}$ to x^*, there exist positive constants β_1 and β_2 such that for k sufficiently large

$$\frac{\beta_1}{\rho\|F_k\|} \le \frac{1}{\|s_k\|} \le \frac{\rho\beta_2}{\|F_k\|}, \tag{2.8}$$

5

The proof of the theorem now follows from Theorem 2.1, and (2.8).

$\square$

The assumption in $(ii)'$ concerning the rate of convergence of $\{x_k\}$ can be replaced by the following weaker statement:

The set

$$Q_1{}^*(\{x_k\}) = \{ \text{ limit points of} \{\frac{\|e_{k+1}\|}{\|e_k\|}\}\} \ ,$$

does not contain one and ∞, for at least one norm.

Clearly the set $Q_1{}^*(\{x_k\})$ depends on the norm selected. The largest element of $Q_1{}^*(\{x_k\})$ is the well-known Q_1-factor. For more detail on this issue, see Chapter 9 of Ortega and Rheinboldt [10].

In terms of secant methods the assumption that $\{x_k\}$ converges to x^* Q-linearly, seems not to be restrictive. In fact if the matrices $\{A_k\}$ satisfy a standard bounded deterioration property, as do the well-known secant methods, then in an appropriate norm, $x_k \rightarrow x^*$, Q-linear. (see Chapter 8 of Dennis and Schnabel [4] for more detail).

Theorem 2.2 tells us that in order to obtain Q-superlinear convergence we should have $r_k = o(\|F_k\|)$ and $\alpha_k \rightarrow 1$. We find it interesting that this is exactly the condition given by Dembo, Eisenstat, and Steihaug [2] for Q-superlinear convergence of their inexact Newton method. Actually, they chose $\alpha_k = 1$ for all k. An obvious choice for the perturbation vector is $r_k = \sigma_k \|F_k\|$ where $\sigma_k \in (0,1]$ and $\sigma_k \rightarrow 0$ as $k \rightarrow \infty$.

3 Primal-dual quasi-Newton interior-point method.

In this section we formulate a primal-dual quasi-Newton interior-point method for solving the constrained optimization problem.

$$\begin{array}{rll} \text{minimize} & f(x) & \\ \text{subject to} & h(x) &= 0 \\ & x &\geq 0 \end{array} \tag{3.1}$$

where $f : \mathbf{R}^n \rightarrow \mathbf{R}$ and $h : \mathbf{R}^n \rightarrow \mathbf{R}^m$ are twice continuously differentiable functions.

The Lagrangian function associated with problem (3.1) is given by

$$l(x,y,z) = f(x) + y^T h(x) - z^T x \tag{3.2}$$

where $y \in \mathbf{R}^m$, and $z \in R^n$ are the Lagrange multipliers associated with the constraints $h(x) = 0$, and $x \geq 0$ respectively.

6

The Karush-Kuhn-Tucker (KKT) conditions for problem (3.1) are

$$F(x,y,z) = \begin{pmatrix} \nabla_x l(x,y,z) \\ h(x) \\ XZe \end{pmatrix} = 0, \quad (x,z) \geq 0, \tag{3.3}$$

where $X = diag(x)$, $Z = diag(z)$ and $e \in \mathbf{R}^n$ is the vector of all ones.

Observe that the inequality constraints in problem (3.1), $x_i \geq 0$, $i = 1, ..., n$, can be written $e_i^T x \geq 0$, $i = 1, ..., n$ where e_i is the $i - th$ natural basis vector, i.e., the $i - th$ component is one while all other components are zero. For x, a feasible point of problem (3.1), we let $\mathcal{B}(x) = \{i : x_i = 0\}$. As is usual in constrained optimization $\mathcal{B}(x)$ is the set of active or binding inequality constraints. We will have need below to consider the gradient of active constraints. It should be clear that this set will be $\{e_i \in \mathbf{R}^n : i \in \mathcal{B}(x)\}$.

In the study of Newton's method, the standard assumptions for problem (3.1) are

A.1. (Existence) There exists (x^*, y^*, z^*) a solution to problem (3.1) and its associated Lagrange multipliers satisfying the KKT conditions (3.3).

A.2. (Smoothness) The Hessian operators $\nabla^2 f, \nabla^2 h_i$, $i = 1, ..., m$ are locally Lipschitz continuous at x^*.

A.3. (Regularity) The set $\{\nabla h_i(x^*) : i = 1, ..., m\} \bigcup \{e_i : i \in \mathcal{B}(x^*)\}$ is linearly independent.

A.4. (Second-Order Sufficiency) For all $\eta \neq 0$ satisfying $\nabla h_i(x^*)^T \eta = 0$, $i = 1, ..., m$; $e_i^T \eta = 0$, $i \in \mathcal{B}(x^*)$ we have $\eta^T \nabla_x^2 l(x^*, y^*, z^*) \eta > 0$

A.5. (Strict Complementarity) For all i, $z_i^* + x_i^* > 0$.

For a nonnegative parameter μ, the perturbed KKT conditions associated to (3.3) are

$$F_\mu(x,y,z) = \begin{pmatrix} \nabla_x l(x,y,z) \\ h(x) \\ XZe - \mu e \end{pmatrix} = 0, \quad (x,z) \geq 0, \tag{3.4}$$

We describe a primal-dual quasi-Newton interior-point method for solving problem (3.1).

7

Algorithm 1. Let $w_0 = (x_0, y_0, z_0)$ be an initial point satisfaying $(x_0, z_0) > 0$.
For $k = 0, 1, ...,$ until convergence do

Step1. Choose $\sigma_k \in (0, 1]$ and set $\mu_k = \sigma_k R_k$ for some $R_k \in \mathbf{R}$.

Step2. Obtain $\Delta w_k = (\Delta x_k, \Delta y_k, \Delta z_k)^T$ as the solution of the linear system

$$M_k \Delta w_k = -F_{\mu_k}(w_k) \tag{3.5}$$

where

$$M_k = \begin{pmatrix} G_k & \nabla h_k & -I_n \\ \nabla h_k^T & 0 & 0 \\ Z_k & 0 & X_k \end{pmatrix}.$$

Step3. Choose $\tau_k \in (0, 1)$ and set
$\alpha_k = min(1, \tau_k \bar{\alpha}_k)$
$\alpha_{ky} = 1$ or $\alpha_{ky} = \alpha_k$
where

$$\bar{\alpha}_k = min \left\{ \frac{-1}{min(X_k^{-1} \Delta x_k, -1)}, \frac{-1}{min(Z_k^{-1} \Delta z_k, -1)} \right\}.$$

Step4. Update
$w_{k+1} = w_k + \Lambda_k \Delta w_k$
where $\Lambda_k = diag(\alpha_k, ..., \alpha_k, \alpha_{ky}, ..., \alpha_{ky}, \alpha_k, ..., \alpha_k)$
in above the three groups of scalars have n, m, and n members respectively.

The choice for R_k will be in general $\|F(w_k)\|$; however we leave it open to obtain a certain amount of needed flexibility in the statement of our theorems in Section 5.

The choice $G_k = \nabla^2_x l(w_k)$ corresponds to Newton's method. For this choice El-Bakry, Tapia, Tsuchiya, and Zhang [5] established local convergence, superlinear convergence, and quadratic convergence for Algorithm 1 for the appropriate choices of τ_k and R_k. Yamashita [13] considered a somewhat different steplength than that described in Step 3, this choice was based on a particular merit function. He then established a global convergence result for his line-search algorithm. El-Bakry et al [5] also gave a global convergence result for a line-search globalization of their form of Algorithm 1. Observe that the choice of steplength in Step 3, $\alpha_k = \tau_k \bar{\alpha}_k$ and $\tau_k \in (0, 1)$ keep x_{k+1} and z_{k+1} positive. If τ_k was chosen to be equal to one, then at least one component of x_{k+1} or z_{k+1} would be zero. We could use different steplengths also for the x and z variables, The obvious choice would be to let $\alpha_{kx} = min(1, \tau_k \bar{\alpha}_{kx})$, and $\alpha_{kz} = min(1, \tau_k \bar{\alpha}_{kz})$, where

$$\bar{\alpha}_{kx} = \frac{-1}{min(X_k^{-1} \Delta x_k, -1)},$$

8

and
$$\bar{\alpha}_{kz} = \frac{-1}{min(Z_k^{-1}\Delta z_k, -1)} \, .$$

Since the asymptotic properties of these choices are essentially the same, we will not concern ourselves with other choices of steplength parameters. It should be clear that the algorithmic choices are the choices of τ_k , σ_k , and G_k the approximation to $\nabla^2_x l(w_k)$. Our objective is to characterize Q-superlinear convergence in terms of the algorithmic choices. A straightforward application of Theorem 2.2 would lead to a characterization in terms of all the variables (x, y, z). Such activity would be incomplete since for equality constrained optimization, where the z-variable is not present, the Bogss-Tolle-Wang characterization is in term of the x-variable alone. Effectively, the y-variable can be removed from the problem as demostrated by Stoer and Tapia [11]. Our first initial efforts in the current research attempted to obtain such a characterization for Algorithm 1; however we could not do so without making assumptions which we considered undesirable. Therefore, we turned to attempting a characterization in terms of the (x, z)-variables and were successful. It follows then that in this application the primary variables are x and z, each carries independent information and can not be removed from the problem. In retrospective we find this occurrence fitting and not surprising.

4   An equivalent formulation.

In this section we imitate the approach taken by Stoer and Tapia [11] in deriving the Boggs-Tolle-Wang characterization for equality constrained optimization. Our task is to construct a quasi-Newton method that involves only the (x, z)-variables, is equivalent to Algorithm 1 of Section 3, and has the form of a damped and perturbed quasi-Newton method as described by (2.2). This equivalence will allow us, in Section 5, to apply our characterization Theorem 2.2.

Assumption **A3** allows us to locally, i.e., in a neighborhood of x^*, consider the projection operator
$$P(x) = I - \nabla h(x)[\nabla h(x)^T \nabla h(x)]^{-1} \nabla h(x)^T \, . \tag{4.1}$$

In turn this allows us to consider the nonlinear equation

$$F_0(x, z) = \begin{pmatrix} P(x)(\nabla f(x) - z) + \nabla h(x)h(x) \\ XZe \end{pmatrix} = 0. \tag{4.2}$$

Observe that $F_0 : \mathbf{R}^{2n} \to \mathbf{R}^{2n}$. We now demonstrate that Algorithm 1 is equivalent to a damped and perturbed quasi-Newton method applied to equation (4.2). Toward this end let (x_k, y_k, z_k), G_k, and μ_k be as in the k-th iteration of Algorithm 1 and consider the linear

9

system

$$\begin{pmatrix} P_k G_k + \nabla h_k \nabla h_k^T & -P_k \\ Z_k & X_k \end{pmatrix} \begin{pmatrix} \Delta x_k \\ \Delta z_k \end{pmatrix} = -(F_0(x_k, z_k) - \mu_k \hat{e}). \tag{4.3}$$

In (4.3), $\hat{e}$ is the $2n$-vector whose first n components are zero and whose last n components are one. We will also need to consider the formula

$$y_k^+ = -(\nabla h_k^T \nabla h_k)^{-1} \nabla h_k^T (G_k \Delta x_k + \nabla f_k - (z_k + \Delta z_k)), \tag{4.4}$$

where $(\Delta x_k, \Delta z_k)$ is the solution of (4.3).

*Proposition 4.1. Let (x^*, y^*, z^*) be a solution of the KKT conditions (3.3) at which the standard assumptions **A1-A5** hold. Then (x^*, z^*) is a solution of the nonlinear equation (4.2) and the standard Newton's method assumptions **S1-S3** hold for F_0 at this solution. Moreover, if $(\Delta x_k, \Delta y_k, \Delta z_k)$ is a solution of the linear system (3.5), then $(\Delta x_k, \Delta z_k)$ is a solution of the linear system (4.3). Conversely, if $(\Delta x_k, \Delta z_k)$ is a solution of the linear system (4.3) and we let $\Delta y_k = y_k^+ - y_k$, where y_k^+ is given by (4.4), then $(\Delta x_k, \Delta y_k, \Delta z_k)$ is a solution of the linear system (3.5).*

Proof. We begin by establishing the equivalence between the linear systems (3.5) and (4.3). Writing out (3.5) in detail gives

$$\begin{array}{lllll} G_k \Delta x_k & + & \nabla h_k \Delta y_k & - & \Delta z_k & = & -(\nabla f_k + \nabla h_k y_k - z_k) \\ \nabla h_k^T \Delta x_k & & & & & = & -h_k \\ Z_k \Delta x_k & & & + & X_k \Delta z_k & = & -X_k Z_k e + \mu_k e \,. \end{array} \tag{4.5}$$

Writing out (4.3) in detail gives

$$\begin{array}{lllll} (P_k G_k + \nabla h_k \nabla h_k^T) \Delta x_k & - & P_k \Delta z_k & = & -(P_k(\nabla f_k - z_k) + \nabla h_k h_k) \\ Z_k \Delta x_k & + & X_k \Delta z_k & = & -X_k Z_k + \mu_k e \,. \end{array} \tag{4.6}$$

We observe that we can write

$$P_k[G_k \Delta x_k + \nabla f_k - (z_k + \Delta z_k)] = G_k \Delta x_k + \nabla f_k - (z_k + \Delta z_k) + \nabla h_k y_k^+ \tag{4.7}$$

where y_k^+ is given by (4.4).

Now, suppose $(\Delta x_k, \Delta y_k, \Delta z_k)$ solves (4.5). Multiplying the first equation by P_k, the second equation by ∇h_k, adding the two resulting equations, and recalling that $P_k \nabla h_k = 0$ leads us to the first equation in (4.6). Hence $(\Delta x_k, \Delta z_k)$ solves (4.6). Conversely, suppose $(\Delta x_k, \Delta z_k)$ solves (4.6). Multiplying the first equation by ∇h_k^T gives the second equation in (4.5). This in turn tells us that the first equation in (4.6) now implies that the left-hand side of (4.7) is zero. Hence the right-hand side is zero and the first equation in (4.5) holds with $y_k + \Delta y_k = y_k^+$. This establishes the equivalence of the two linear systems (4.5)

10

and (4.6).

If (x^*, y^*, z^*) solves (3.3), then clearly (x^*, z^*) solves (4.2). Observing that $P(x)(\nabla f(x) - z) = P(x)(\nabla f(x) + \nabla h(x)y^+(x^*, z^*) - z)$ and $y^+(x^*, z^*) = y^*$ we see that

$$F_0'(x^*, z^*) = \begin{pmatrix} P_*\nabla_x^2 l(x^*, y^*, z^*) + \nabla h_* \nabla h_*^T & -P_* \\ Z_* & X_* \end{pmatrix}. \tag{4.8}$$

An argument along the lines of the one given above can be used to show that the linear system

$$F_0'(x^*, z^*) \begin{pmatrix} \eta_x \\ \eta_z \end{pmatrix} = 0 \tag{4.9}$$

is equivalent to the linear system

$$F'(x^*, y^*, z^*) \begin{pmatrix} \eta_x \\ \eta_y \\ \eta_z \end{pmatrix} = 0 \tag{4.10}$$

where F is given by (3.3). Under the standard assumptions **A1-A5**, for F given by (3.3), we know that $F'(x^*, y^*, z^*)$ is nonsingular. Hence $F_0'(x^*, z^*)$ must also be nonsingular. It should be clear that F_0 and F have the same smoothess properties. This says that assumptions **S1-S3**, appropriately stated, hold for F_0 at (x^*, z^*). We have now established our equivalence proposition.

$\square$

We have shown that obtaining (x_k, z_k) from Algorithm 1 can be viewed as obtaining (x_k, z_k) from a damped and perturbed quasi-Newton method applied to the nonlinear equation $F_0(x, z) = 0$ given by (4.2). Moreover, the approximate Jacobian has the form

$$\begin{pmatrix} P_k G_k + \nabla h_k \nabla h_k^T & -P_k \\ Z_k & X_k \end{pmatrix} \tag{4.11}$$

and the Jacobian at the solution is given by (4.8).

We are now ready to state our Q-superlinear convergence results.

5   Q-superlinear convergence characterization.

In this section we apply the theory developed in Section 2 to the primal-dual quasi-Newton interior-point method described by Algorithm 1 of Section 3. Recall that G_k is our approximation to $G_* = \nabla^2 f(x^*) + \nabla^2 h(x^*)y^*$. Also R_k appears in Step 1 of Algorithm 1.

Theorem 5.1. *Let $\{(x_k, y_k, z_k)\}$ be generated by Algorithm 1. Assume that $\{(x_k, y_k, z_k)\}$*

11

converges to (x^*, y^*, z^*) and assumptions **A1-A5** *hold at* (x^*, y^*, z^*). *Furthermore, assume that τ_k and σ_k have been chosen so that*

(i) $\tau_k \to 1$.

(ii) $\sigma_k \to 0$.

Assume that either $R_k = O(\|s_k\|)$, where $s_k = (x_{k+1}, y_{k+1}, z_{k+1}) - (x_k, y_k, z_k)$, or $R_k = O(\|F(x_k, y_k, z_k)\|)$ and $\{(x_k, y_k, z_k)\}$ converges to (x^*, y^*, z^*) Q-linearly. Then $\{(x_k, y_k, z_k)\}$ converges Q-superlinearly to (x^*, y^*, z^*) if and only if

$$\frac{\|(G_k - G_*)(x_{k+1} - x_k)\|}{\|x_{k+1} - x_k\| + \|y_{k+1} - y_k\| + \|z_{k+1} - z_k\|} \to 0. \tag{5.12}$$

Assume that either $R_k = O(\|s_k\|)$ where $s_k = (x_{k+1}, z_{k+1}) - (x_k, z_k)$ or $R_k = O(\|F_0(x_k, z_k)\|)$, where F_0 is given by (4.2), and $\{(x_k, z_k)\}$ converges to (x^*, z^*) Q-linearly Then $\{(x_k, z_k)\}$ converges Q-superlinearly to (x^*, z^*) if and only if

$$\frac{\|P_k(G_k - G_*)(x_{k+1} - x_k)\|}{\|x_{k+1} - x_k\| + \|z_{k+1} - z_k\|} \to 0. \tag{5.13}$$

Proof. The proof of the theorem follows by applying Theorem 2.1, Theorem 2.2, and Proposition 4.1, and using (3.5), (4.8), and (4.11). We have used the following fact concerning norms in finite dimensional spaces. Let $u \in \mathbf{R}^n$ and $v \in \mathbf{R}^m$. Also let $\| \ \|_n$ be a norm on $\mathbf{R}^n$, $\| \ \|_m$ a norm on $\mathbf{R}^m$, and $\| \ \|_{n+m}$ a norm on $\mathbf{R}^{n+m}$. Then there exist positive constants θ_1 and θ_2 such that

$$\theta_1(\|u\|_n + \|v\|_m) \le \|(u, v)\|_{n+m} \le \theta_2(\|u\|_n + \|v\|_m). \tag{5.14}$$

A proof of (5.14) can be obtained by working with the l_1 norm and the equivalence of norms property. We also used the fact that $\tau_k \to 1$ implies $\alpha_k \to 1$ (see Step 3 of Algorithm 1) under our assumptions. This fact can be found in Yamashita and Yabe [14]. Finally, we have removed all quantities that converged to zero and were redundant in the characterization result. $\square$

Yamashita and Yabe [14] gave a characterization which has the flavor of (5.12). However, their assumptions were somewhat more restrictive.

12

References

[1] P. T. Boggs, J. W. Tolle, and P. Wang, *On the local convergence of quasi-Newton methods for constrained optimization*, SIAM J. Control Optim., 20 (1982), pp. 161-171.

[2] R. S. Dembo, S. C. Eisenstat and T. Steihaug, *Inexact Newton Methods*, SIAM J. Numer. Anal., 19 (1982), pp. 400-408.

[3] J. E. Dennis, Jr. and J. J. Moré, *A characterization of superlinear convergence and its application to quasi-Newton methods* Math. Comp., 28 (1974), pp. 549-560.

[4] J. E. Dennis, Jr. and R. B. Schnabel, *Numerical Methods for Unconstrained Optimization and Nonlinear Equations*, (1983), Prentice-Hall, Englewood Cliffs, NJ.

[5] A. S. El-Bakry, R. A. Tapia, T. Tsuchiya, and Y. Zhang, *On the formulation of the primal-dual interior point method for nonlinear programming*, Technical Report 92-40, Department of Computational and Applied Mathematics, Rice University, Houston, Tx.

[6] R. Fontecilla, T. Steihaug and R. A. Tapia, *A convergence theory for a class of quasi-Newton methods for constrained optimization*, SIAM J. Numer. Anal., 24, (1987), pp. 1133-1151.

[7] S. T. Glad, *Properties of updating methods for the multipliers in augmented Lagrangians*, J. Optim. Theory Appl., 28 (1979), pp. 135-156.

[8] S. P. Han, *Superlinearly convergent variable metric algorithms for general nonlinear programming problems*, Math. Programming, 11 (1976), pp. 263-282.

[9] J. Nocedal and M. Overton, *Projected Hessian updating algorithms for nonlinear constrained optimization*, SIAM J. Numer. Anal. 22 (1985) pp. 821-850.

[10] J. M. Ortega and W. C. Rheinboldt, *Iterative Solutions of Nonlinear Equations in Several Variables*, (1970), Academic Press, New York.

[11] J. Stoer and R. A. Tapia. *On the characterization of q-superlinear convergence of quasi-Newton methods for constrained optimization*, Mathematics of Computation, 49 (1987), pp. 581-584.

[12] R. A. Tapia, *Diagonalized multiplier methods and quasi-Newton methods for constrained optimization*, J. Optim. Theory Appl., 22 (1977), pp. 135-194.

[13] H. Yamashita, *A globally convergent primal-dual interior point method for constrained optimization*, Technical Report, Mathematical System Institute, Inc.. Japan, 1992.

[14] H. Yamashita and H. Yabe *superlinear and quadratic convergence of primal-dual interior point methods*, Technical Report, Mathematical System Institute, Inc.. Japan, 1993.